Noname manuscript No.
(will be inserted by the editor)

Supplementary Material for "Estimation of Treatment Effects and Model Diagnostics with Two-way Time-Varying Treatment Switching: an Application to a Head and Neck Study"

QINGXIA CHEN · FAN ZHANG · MING-HUI CHEN · XIUYU JULIE CONG

Received: date / Accepted: date

1 Regularity Assumptions

Assumption 1. Treatment R is completely randomized and $T_D^*(a) = T_D(a)$ if a subject never changes treatment.
Assumption 2. Given $(R = 0, \mathbf{Z}, U = 1, T_P = s)$, that is, a subject in the control arm has disease progression at time s and covariates $\mathbf{Z}$, or $(R = 1, \mathbf{Z}, U = 1, T_P = s)$, $N_V(s+t), t \geq 0$, is independent of the potential outcomes $\{T_D^*(0), T_D^*(1)\}$.
Assumption 3. Given $(R = 0, \mathbf{X}, U = 0)$, that is, a subject in the control arm has baseline covariates $\mathbf{X}$ and is progression free before death, or $(R = 1, \mathbf{X}, U = 0)$, $N_V(t)$ is independent of the potential outcomes $\{T_D^*(0), T_D^*(1)\}$.
Assumption 4. The censoring time is independent of T_D, T_G, and T_P given the observed covariates.
Assumption 5. For progression subjects, T_P is independent of $\mathbf{Z}$ given R and $\mathbf{X}$.
Assumption 6. The true parameters values of the $\boldsymbol{\beta}$'s, $\boldsymbol{\gamma}$'s and $\boldsymbol{\alpha}$'s, still denoted as $\Phi \equiv (\boldsymbol{\beta}_0^T, \beta_1, \boldsymbol{\beta}_2^T, \boldsymbol{\gamma}_0^T, \boldsymbol{\gamma}_1^T, \boldsymbol{\gamma}_2^T, \boldsymbol{\alpha}^T)^T$, belong to a bounded set in real

QINGXIA CHEN
Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN 37232
E-mail: cindy.chen@vumc.org

FAN ZHANG *
Pfizer Inc., Groton, CT, 06340
* This work was done when Dr. Zhang was at University of Connecticut.

MING-HUI CHEN
Department of Statistics, University of Connecticut, 215 Glenbrook Road, U-4120, Storrs, CT 06269

XIUYU JULIE CONG**
Everest Medicines, Shanghai, China.
** This work was done when Dr. Cong was at Boehringer Ingelheim Pharmaceuticals Inc., 900 Ridgebury Rd, Ridgefield, CT, 06811, USA.

Euclidean space. Moreover, the true cumulative baseline functions (H_0, H_1, H_2) are continuously differentiable in $[0, \tau]$ with $H_k'(t) > 0$, $k = 0, 1, 2$, where τ is the study duration.

ASSUMPTION 7. If there is some constant ν such that $\nu^T(1, R, \mathbf{Z}) = 0$ with probability one, then $\nu = 0$. Additionally, we assume $(R, \mathbf{Z})$ to have a bounded support and there exits a continuous component of $\mathbf{X}$ such that its coefficient in model (1) is nonzero.

ASSUMPTION 8. With probability one, $P(C \geq \tau | R, \mathbf{Z}) > 0$ and $P(V = 1 | R, \mathbf{Z}, T_P) \in (\mu_0, \mu_1)$ for some constant $0 < \mu_0 < \mu_1 < 1$.

2 Proof of Theorem 1–3

Proof of Theorem 1. Let $l_n(\Phi, H_1, H_2, H_3)$ denote the observed log-likelihood function for (Φ, H) and $H\{t\} = \{H(t) - H(t-)\}$. First, it is easy to see that if $\widehat{H}_k\{t\} = \infty$, then $l_n(\widehat{\Phi}, \widehat{H}_1, \widehat{H}_2, \widehat{H}_3) = -\infty$. Moreover, this also holds if the jump size of $\widehat{H}_k$ at the corresponding events is zero. Thus, the jump sizes of $\widehat{H}_k$ at the corresponding events are positive and finite so the derivatives of $l_n(\Phi, H_1, H_2, H_3)$ with respect to each jump size of $\widehat{H}_k$ should be zero at $(\widehat{\Phi}, \widehat{H}_1, \widehat{H}_2, \widehat{H}_3)$. This gives

$$\widehat{H}_0\{Y_i\} = \frac{\sum_{G_j=1} I(Y_j = Y_i)}{\left\{\sum_{G_j=1, Y_j \geq Y_i} e^{\hat{\eta}_{Dj}(Y_i)} + \sum_{G_j=4, Y_j \geq Y_i} \frac{(1-\hat{p}_{uj})\hat{S}_{Dj}(Y_j) e^{\hat{\eta}_{Dj}(Y_i)}}{(1-\hat{p}_{uj})\hat{S}_{Dj}(Y_j) + \hat{p}_{uj}\hat{S}_{Uj}(Y_j)}\right\}} \tag{1}$$

$$\widehat{H}_1\{W_i\} = \frac{\sum_{G_j \in \{2,3\}} I(W_j = W_i)}{\left\{\sum_{G_j \in \{2,3\}, W_j \geq W_i} e^{\hat{\eta}_{Pj}(W_i)} + \sum_{G_j=4, Y_j \geq W_i} \frac{\hat{p}_{uj}\hat{S}_{Pj}(Y_j) e^{\hat{\eta}_{Pj}(W_i)}}{(1-\hat{p}_{uj})\hat{S}_{Dj}(Y_j) + \hat{p}_{uj}\hat{S}_{Pj}(Y_j)}\right\}} \tag{2}$$

$$\widehat{H}_2\{(Y-W)_i\} = \frac{\sum_{G_j=1} I((Y-W)_j = (Y-W)_i)}{\left\{\sum_{G_j \in \{2,3\}, (Y-W)_j \geq (Y-W)_i} e^{\hat{\eta}_{Gj}((Y-W)_i)}\right\}} \tag{3}$$

where $\hat{\eta}_{Dj}(t) = \hat{\beta}_{01} R_j + \hat{\beta}_{02} V_j(t) + \hat{\beta}_{03} V_j(t) R_j + \hat{\gamma}_0^T \mathbf{X}_j$, $\hat{\eta}_{Pj}(t) = \hat{\beta}_1 R_j + \hat{\gamma}_1^T \mathbf{X}_j$, $\hat{\eta}_{Gj}(t) = \hat{\beta}_{21} R_j + \hat{\beta}_{22} V_j(t + T_P) + \hat{\beta}_{23} V_j(t + T_P) R_j + \hat{\gamma}_2^T (\mathbf{Z}_j, W_j)$. In addition, we let $\hat{S}_{Dj}(t) = \exp\{-\int_0^t e^{\hat{\eta}_{Dj}(s)} d\hat{H}_0(s)\}$, $\hat{S}_{Pj}(t) = \exp\{-\int_0^t e^{\hat{\eta}_{Pj}(s)} d\hat{H}_1(s)\}$, and $\hat{p}_{uj} = \widehat{P}(U = 1 | R_j, \mathbf{X}_j)$. Equation (1) implies $\widehat{H}_0\{s\} \leq \sum_{G_j=1} I(Y_j = s) / \sum_{G_j=1, Y_j \geq s} c_0$, where c_0 is a positive lower bound of $e^{\hat{\eta}_{Dj}(t)}$. Since

$$\frac{1}{n} \sum_{G_j=1, Y_j \geq s} 1 \to E[I(U = 0, \cdot \leq C, \cdot \geq s)] > 0,$$

we obtain

$$\limsup_n \widehat{H}_0(\tau) \le \limsup_n \frac{n^{-1}\sum_{i=1}^n \sum_{j=1}^n I(G_j = 1, Y_j = Y_i)}{c_0 n^{-1}\sum_{G_j=1, Y_j \ge s} 1} < \infty.$$

Similarly, equations (2) and (3) yield that $\limsup_n \widehat{H}_1(\tau)$ and $\limsup_n \widehat{H}_2(\tau)$ are both finite.

By Helly's selection theorem, for any subsequence, we can choose a further subsequence such that $\widehat{H}_k$ weakly converges to an increasing function H_k^* for $k = 1, 2, 3$. Moreover, we can assume $\widehat{\theta} \to \theta^*$. We then show $H_k^* = H_k$ and $\Phi^* = \Phi$. To this end, we construct $\tilde{H}_k$ such that $\tilde{H}_k$ has jumps at the same events as $\widehat{H}_k$; moreover, the jumps of $\tilde{H}_k$ is the right-hand side of (1) to (3) except that the parameters on the right-hand side are set to be the true values. It is straightforward to verify that $\tilde{H}_k$ converges uniformly to the true function H_k. Furthermore, we can show that $d\widehat{H}_k/d\tilde{H}_k$ converges uniformly to dH_k^*/dH_k.

Since $l_n(\widehat{\Phi}, \widehat{H}_1, \widehat{H}_2, \widehat{H}_3) - l_n(\Phi, \tilde{H}_1, \tilde{H}_2, \tilde{H}_3) \ge 0$, we take limits and then expectation on both sides, which leads to the conclusion that the Kullback–Leibler information between $(\theta^*, H_1^*, H_2^*, H_3^*)$ and (θ, H_1, H_2, H_3) is non-positive. As the Kullback–Leibler information is always non-negative, this immediately implies that the log-likelihood function at $(\theta^*, H_1^*, H_2^*, H_3^*)$ is equal to the log-likelihood function at (θ, H_1, H_2, H_3) with probability one. Thus, this equality holds for all subjects in Groups 1 to 4 as defined in Section 3.2. Comparing the differences of the log-likelihood functions from subjects in Group 2 and Group 3, we have

$$\begin{aligned}
&(H_2^*)'(G)e^{\beta_{21}^* R + \beta_{22}^* V(Y) + \beta_{23}^* V(Y)R + \gamma_2^{*T}(\mathbf{Z}^T, W)^T} \\
&= (H_2)'(G)e^{\beta_{21} R + \beta_{22} V(Y) + \beta_{23} V(Y)R + \gamma_2{}^T(\mathbf{Z}^T, W)^T},
\end{aligned}$$

so by Assumptions 7 and 8, $H_2^* = H_2, \beta_{21}^* = \beta_{21}, \beta_{22}^* = \beta_{22}$, $\beta_{23}^* = \beta_{23}$ and $\gamma_2^* = \gamma_2$. Now in the log-likelihood for subjects in Group 1, we let $Y = 0$ and obtain

$$\frac{(H_0^*)'(0)e^{\beta_{01}^* R + \gamma_0^{*T}\mathbf{X}}}{1 + e^{\alpha_0^* + \alpha_1^* R + \alpha_2^{*T}\mathbf{X}}} = \frac{(H_0)'(0)e^{\beta_{01} R + \gamma_0{}^T\mathbf{X}}}{1 + e^{\alpha_0 + \alpha_1 R + \alpha_2{}^T\mathbf{X}}}.$$

Similarly, in the log-likelihood for subjects in Group 3, we let $W = 0$ and $Y = 0$ and obtain

$$\frac{(H_1^*)'(0)e^{\beta_1^* R + \gamma_1^{*T}\mathbf{X}} e^{\alpha_0^* + \alpha_1^* R + \alpha_2^{*T}\mathbf{X}}}{1 + e^{\alpha_0^* + \alpha_1^* R + \alpha_2^{*T}\mathbf{X}}} = \frac{(H_1)'(0)e^{\beta_1 R + \gamma_1{}^T\mathbf{X}} e^{\alpha_0 + \alpha_1 R + \alpha_2{}^T\mathbf{X}}}{1 + e^{\alpha_0 + \alpha_1 R + \alpha_2{}^T\mathbf{X}}}.$$

Comparing the above equations, so $\alpha_1^* = \alpha_1, \alpha_2^* = \alpha_2$. Since one component of $\mathbf{X}$ is continuous and has non-zero coefficient in α_2, the above equation gives $\alpha_0^* = \alpha_0$. Finally, after integrating the likelihood equality function for Group 2 for W from 0 to Y, we have

$$\frac{[1 - \exp\{-H_1^*(Y)e^{\beta_1^* R + \gamma_1^{*T}\mathbf{X}}\}]e^{\alpha_0^* + \alpha_1^* R + \alpha_2^{*T}\mathbf{X}}}{1 + e^{\alpha_0^* + \alpha_1^* R + \alpha_2^{*T}\mathbf{X}}}$$

$$= \frac{[1-\exp\{-H_1(Y)e^{\beta_0 R+\gamma_1{}^T\mathbf{X}}\}]e^{\alpha_0^*+\alpha_1^* R+\alpha_2^{*T}\mathbf{X}}}{1+e^{\alpha_0+\alpha_1+\alpha_2{}^T\mathbf{X}}}.$$

Thus, $H_1^* = H_1$ and $\beta_1^* = \beta_1, \gamma_1^* = \gamma_1$. On the other hand, integrating the likelihood equality function for subjects in Group 1 for Y from 0 to Y gives

$$\frac{1-\exp\{-\int_0^Y e^{\beta_{01}^* R+\beta_{02}^* V(s)+\beta_{03}^* V(s)R+\gamma_0^{*T}\mathbf{X}}dH_0^*(s)\}}{1+e^{\alpha_0^*+\alpha_1^* R+\alpha_2^{*T}\mathbf{X}}}$$
$$= \frac{1-\exp\{-\int_0^Y e^{\beta_{01} R+\beta_{02} V(s)+\beta_{03} V(s)R+\gamma_0{}^T\mathbf{X}}dH_0^*(s)\}}{1+e^{\alpha_0+\alpha_1 R+\alpha_2{}^T\mathbf{X}}}$$

so $\beta_0^* = \beta_0, \gamma_0^* = \gamma_0$ and $H_0^* = H_0$.

We have proved that $\widehat{\theta} \to \theta$ and $\widehat{H}_k$ converges weakly to H_k. The latter can be further strengthened to uniform convergence in $[0,\tau]$ since H_k is continuous. Therefore, Theorem 1 holds.

Proof of Theorem 2. The proof of Theorem 2 follows from the same argument in proving Theorem 2 in [2]. In particular, Assumptions 1-4 and 6 in Appendix hold for our specific models. Their first identifiability condition (C.5) has been verified in the proof of Theorem 1. To complete the proof, we only need to verify the second identifiability of their condition (C.7). Consider the score function along a sub model $H_k + \epsilon \int f_k dH_k$ and $\Phi + \epsilon\nu$ where $\nu = (\boldsymbol{\beta}_0, \boldsymbol{\gamma}_0, \beta_1, \boldsymbol{\gamma}_1, \boldsymbol{\beta}_2, \boldsymbol{\gamma}_2, \boldsymbol{\alpha})$. If this score function is zero with probability one, then we need to show that $f_k = 0$ and $\nu = 0$. For subjects in Group 2, the score equation is

$$\begin{aligned} 0 = \quad & f_1(W) + \eta_P - \int_0^W f_1(t)e^{\eta_P}dH_1(t) - H_1(W)e^{\eta_P}\eta_P + f_2(G) + \eta_G(G) \\ & - \int_0^G f_2(t)e^{\eta_G(t)}dH_2(t) - \int_0^G e^{\eta_G(t)}\eta_G(t)dH_2(t) + \frac{e^{\alpha_0+\alpha_1 R+\alpha_2^T\mathbf{X}}(\xi_0+\xi_1 R+\xi_2^T\mathbf{X})}{(1+e^{\alpha_0+\alpha_1 R+\alpha_2^T\mathbf{X}})^2}. \end{aligned}$$

For subjects in Group 3, we obtain the score equation to be

$$\begin{aligned} 0 = & f_1(W) + \eta_P - \int_0^W f_1(t)e^{\eta_P}dH_1(t) - H_1(W)e^{\eta_P}\eta_P - \int_0^G f_2(t)e^{\eta_G(t)}dH_2(t) \\ & - \int_0^G e^{\eta_G(t)}\eta_G(t)dH_2(t) + \frac{e^{\alpha_0+\alpha_1 R+\alpha_2^T\mathbf{X}}(\xi_0+\xi_1 R+\xi_2^T\mathbf{X})}{(1+e^{\alpha_0+\alpha_1 R+\alpha_2^T\mathbf{X}})^2}. \end{aligned} \tag{4}$$

The difference between (4) and (4) gives $f_2(G)+\eta_G(G) = 0$, so by Assumption 7, $f_2 = 0, \boldsymbol{\beta}_2 = 0$ and $\boldsymbol{\gamma}_2 = 0$.

Using this result and equation (4), the score equation for subjects in Group 4 becomes

$$\int_0^Y f_0(t)e^{\eta_D(t)}dH_0(t) + \int_0^Y e^{\eta_D(t)}\eta_D(t)dH_0(t) + \frac{e^{\alpha_0+\alpha_1 R+\alpha_2^T\mathbf{X}}(\alpha_0+\alpha_1 R+\alpha_2^T\mathbf{X})}{(1+e^{\alpha_0+\alpha_1 R+\alpha_2^T\mathbf{X}})^2} = 0. \tag{5}$$

On the other hand, for subjects in Group 1,

$$f_0(Y) + \eta_D(Y) - \int_0^Y f_0(t)e^{\eta_D(t)}dH_0(t) - \int_0^Y e^{\eta_D(t)}\eta_D(t)dH_0(t)$$
$$-\frac{e^{\alpha_0+\alpha_1 R+\alpha_2^T \mathbf{X}}(\alpha_0 + \alpha_1 R + \alpha_2^T \mathbf{X})}{(1+e^{\alpha_0+\alpha_1 R+\alpha_2^T \mathbf{X}})^2} = 0. \tag{6}$$

Then the difference between (5) and (6) gives $f_0(Y)+\eta_D(Y) = 0$, which further gives $f_0 = 0, \boldsymbol{\beta}_0 = 0$ and $\boldsymbol{\gamma}_0 = 0$. As a result, (6) becomes $\alpha_0+\alpha_1 R+\alpha_2^T \mathbf{X} = 0$ and hence $\boldsymbol{\alpha} = 0$. This further combined with equation (4) gives $f_1 = 0$, $\boldsymbol{\beta}_1 = 0$, and $\boldsymbol{\gamma}_1 = 0$. We have verified condition (C.7) in [2]. According to their results, our Theorem 2 holds.

Moreover, from Theorem 3 in [2], we also conclude that the inverse of the observed information is a consistent estimator for the asymptotic covariance.

Proof of Theorem 3. The proof of Theorem 3 follows the proof of Theorem 3 in [1] so we omit the details here.

References

1. Zeng, D., Chen, Q., Chen, M.H., Ibrahim, J.G., et al.: Estimating treatment effects with treatment switching via semicompeting risks models: an application to a colorectal cancer study. Biometrika **99**(1), 167–184 (2012)
2. Zeng, D., Lin, D.Y.: A general asymptotic theory for maximum likelihood estimation in semiparametric regression models with censored data. Statistica Sinica **20**, 871–910 (2010)